DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES

2001-02

McMASTER UNIVERSITY

Department of Economics
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

http://socserv.socsci.mcmaster.ca/~econ/

Inferring Willingness-to-Pay for Health Attributes of Air Quality using Information on Ranking of Alternatives and Cognitive Ability of Respondents*

R.A. Muller, McMaster University, Department of Economics

A.L. Robb, McMaster University, Department of Economics

A. Diener, University of Nebraska Medical Center, Preventive and Societal Medicine

March 20, 2001

Please check with authors before quoting or citing

Corresponding Author:
Andrew Muller
Department of Economics
McMaster University
Hamilton, Ont. Canada L8S 4M4
e-mail: mullera@mcmaster.ca

Proposed Running Head: WTP with Information on Ranking and Cognition

JEL Classification Numbers: I10, Q26

---

*We are indebted to Vic Adamowicz for the suggestion that led to this paper. The usual disclaimer applies. We acknowledge with thanks the funding for the survey employed in this paper from the Ontrario Ministry of Environment and Energy and the Regional Municipality of Hamilton-Wentworth through the Hamilton Air Quality Initiative.

Inferring Willingness-to-Pay for Health Attributes of Air Quality using Information on Ranking of Alternatives and Cognitive Ability of Respondents

R.A. Muller, A.L. Robb, A. Diener

Abstract

We investigate the use of variable scale parameters to account for heterogeneity in responses to a stated preference survey of willingness to pay for air quality. The survey instrument collected full rankings of four alternatives in each of nine choice sets. Nine of 54 pairwise comparisons involved dominated alternatives; selection of a dominated alternative is interpreted as indicating cognitive difficulties. We allow the scale parameter in a conditional logit model to vary by the ranking considered (first, second or third choice) and by the degree of cognitive difficulty encountered. We investigate the effect of this on the magnitude and precision of estimated willingness-to-pay using empirically-based confidence intervals. The size of the scale parameter is significantly and positively related to the frequency with which respondents chose dominated choices, even after corrections designed to eliminate endogeneity bias. Ignoring heterogeneity in the cognitive ability of respondents biases willingness-to-pay estimates upwards. Consideration of second and third choices narrows the confidence intervals on WTP estimates. Introducing a variable scale parameter related to ranking by itself does not improve the precision of the estimate. Sample STATA code is provided.

Full text at www.socsci.mcmaster.ca/mullera/papers/aqhw05.pdf

Survey Instrument at www.socsci.mcmaster.ca/mullera/papers/aqhw-questionaire.pdf

Inferring Willingness-to-Pay for Health Attributes of Air Quality using Information on Ranking of Alternatives and Cognitive Ability of Respondents

Introduction

Choice experiments featuring conjoint analysis of stated preferences are increasingly being employed to measure willingness-to-pay for change in health or environmental services. These techniques originated in the marketing literature (see Louviere, 1987, 1988). Early extensions to the environmental and health literature include papers by Adamowicz et al. (1994), Schulze et al. (1995), and Ryan et.al. (1998). A choice experiment relies on surveys that require respondents to choose a preferred outcome from a group of two or more alternatives. Each group of alternatives is called a choice set. Each alternative in the set is characterized by a specified level of each of a number of attributes of the service in question together with a form of payment. The attributes of the alternatives are varied randomly or systematically across choice sets and individual respondents. A binomial or multinomial choice model is applied to explain observed choices as a function of the level of each attribute, possibly interacted with demographic or other variables. Marginal willingness-to-pay for any attribute is then computed as the ratio of the coefficient on the attribute to the coefficient on the payment vehicle.

The standard multinomial choice model contains a scale parameter, μ, which is associated with the precision of choice. As μ tends to zero, the probability of choosing the alternative with highest predicted utility approaches unity. As μ tends to infinity the probabilities of all choices tend to equality; that is, the probability distribution of choices becomes uniform. The scale parameter cannot be identified in the basic choice experiment described above, consequently it is usually normalized to unity. The scale parameter plays an important role, however, in extensions to the basic model which are discussed below.

The basic model uses only information about the first ranked choices and generally ignores any other preference information that may be provided by the survey. However, survey data often contain more information than is used by this standard model. For example, subjects may rank three or more alternatives from most to least preferred and efficiency suggests using the additional information conveyed by the second and subsequent rankings. This can be accomplished by combining information about first and subsequent choices. The likelihood of the first choice in any choice set can be computed according to the basic model. The likelihood of the second choice can be computed according to the same model applied to the choices remaining after the preferred option is dropped, and so on. The likelihood of any observed ranking is computed as the product of the likelihood of the individual choices. Parameter estimates are once again obtained by maximizing the likelihood of the sample.[2]

Some surveys collect data both on hypothetical choices (stated preferences) and on actual past choices (revealed preference). Recently considerable attention has been paid to developing methods for combining this data.[3] The key innovation lies in allowing common parameters to appear in the separate multinomial choice models for the stated preference and revealed preference questions. The likelihood of any individuals' response is the product of the likelihood of the stated choice and the revealed choice and the parameters can be estimated by maximizing the joint likelihood.

---

[2]See Chapman and Stalin (1982) and Schulze (1996) for early applications in marketing and environmental economics respectively. Laughton (2000) also exploits ranking data.

[3]See Henscher et al., 1999, for a review and exposition.

In combining revealed and stated preference data it is natural to allow the scale parameter of the multinomial choice model to differ according to the source of preference data and in fact the distinction between the two data sources identifies the ratio of the two parameters. Allowing the scale parameter to vary across sources of preference data generally improves the precision of the estimated parameters.

Recently, analysts of choice experiments have given increasing attention to heterogeneity across subjects or groups. By this we do not mean differences across demographic categories, which can be incorporated by interacting demographic variables with health or environmental attributes, but rather the possibility that individuals differ systematically in their ability to answer survey questions. This difficulty can be addressed by allowing the scale parameter to vary across different groups of individuals. For example, Johnson et al (2000) allow the scale parameter to be a function of demographic variables, including score on a test of understanding built into the survey.

Because allowing for heterogeneity in the scale parameter is a relatively new development, we believe it is useful to explore its value in analysing a variety of data sets. In this paper we examine the success of allowing for heterogeneity in inferring willingness to pay for four attributes of air quality in the Regional Municipality of Hamilton-Wentworth. The survey instrument is particularly interesting for this purpose, because it features large choice sets with four alternatives. Respondents were asked to rank all four alternatives, so we can combine information on first, second and third preferences. It is natural to allow the scale parameter to vary across these data sources. In addition, the partial factorial design employed yielded 9 pairwise comparisons (out of 54) in which one choice was strongly dominated by another. We interpret a subjects' choice of a dominated alternative (over a dominated one) to be evidence of cognitive difficulties with the survey instrument. We group subjects according to cognitive level and allow the error parameter to vary accordingly. Our chief interest lies in obtaining stable and credible inferences about willingness-to-pay from the survey data. Since willingness-to-pay is imputed as the ratio of two coefficients, exact confidence intervals on willingness to pay are not available. We derive empirical standard errors using the Robb-Krinsky (1991) method.

Data and Methods

The Survey

The data in this paper were collected in a mail survey of randomly selected households[4] in the Regional Municipality of Hamilton Wentworth.[5] The survey was conducted in February of 1997 and sent to 1908 households. Some 259 of the surveys were undeliverable leaving 1649 surveys of which 515 were returned for a response rate of 31%. The usable sample was smaller than this as a consequence of incomplete responses.[6] As reported in Diener (1999) the sample reasonably represents the underlying population of the Hamilton Wentworth region when compared to 1991 Census data. We do not spend time on the representativeness issue here as the focus of this paper is primarily methodological.

In addition to collecting basic demographic and opinion data, the survey asked individuals

---

[4] Local tax records were used to draw the sample. See Diener (1999) for more details.

[5] Funding was provided by the Ontario Ministry of the Environment and the Regional Municipality of Hamilton-Wentworth through the Hamilton-Wentworth Air Quality Initiative. More details of the background and the methodology are reported in Diener (1999).

[6] A reward was offered for returning the response in terms of participation in a draw. This encouraged some individuals to fill in a minimum amount of information and return the survey.

to rank nine groups of four states of the world. Each state was characterized by different amounts of air quality attributes, health and taxes. The sample question used to introduce respondents to the ranking process is shown as Figure 1. The bold characters in the last row of the table are a sample ranking of the alternatives and it is this row that respondents would fill in.

Each respondent was asked to provide rankings for nine sets of alternatives. In each of the nine choice sets, Choice A was always the same and the attribute levels in this option were chosen to represent the status quo in Hamilton (as explained more fully in the complete survey document). The other choices were formed by varying the three air quality attributes and the health effects by one-third (better or worse) or by leaving them the same, and by varying the monthly property taxes (up or down) or by leaving them the same.

The options were chosen based on a partial factorial design, which was picked to allow for high order interaction effects to be identified in the responses.[7] We do not explore any such interaction effects in this paper, however. There are five 'attributes"each having three possible levels. Thus there are $(3)^5$ or 243 alternatives that can be compared to Choice A. The partial factorial design selected 27 (one-ninth) of these, which we arranged in 9 choice sets of the sort shown in Figure 1.

Finally we created 6 different versions of each choice set by allowing for 6 different tax increments ($5, $10, $15, $20, $25, and $50) and we formed these into 6 different versions of the questionnaire (each with 9 choice sets) in such a way that each choice set was presented with each tax treatment and each respondent faced each of the tax increases at least once.

This procedure for generating choice sets produced some choices which were dominated

[7] See Petersen (1985) for a good discussion of factorial design.

by others in the same set, in the sense that the level of at least one of its attributes was inferior to another choice and none were better. (We assumed that improvements in air quality attributes and reductions in taxes would always be preferred). As a result, 9 of the 54 possible pairwise comparisons in the nine choice sets involve a dominance relationship.[8] It is the failure to rank these dominance relationships correctly that allows us to identify respondents who are less able to conduct the ranking exercise and to explore the difference it makes on the estimates when these less able subjects are included in the estimation. We will denote a case in which a respondent's ranking indicates that a particular choice is preferred to a choice which dominates it as a "dominance violation".

Statistical Model

Choice experiments are analysed with a simple random utility model. Let the choices presented to the i-th subject be grouped into N choice sets, each comprising J alternatives. Express the utility derived by the i-th individual i from the j-th alternative of the n-th choice set as the sum of a systematic component, V_{inj} , and a random component, e_{inj}. Let the systematic component be a linear function of the vector of attributes of this alternative, i.e. $V_j = X_j' \beta$.

Suppressing the subscripts i and n for convenience, we have

$$U_j = X_j' \beta + \epsilon_j \qquad (1)$$

[8] A close look at Figure 1 will reveal that the choices in this sample questionnaire are such that C is preferred to A which in turn is preferred to D which is preferred to B. Of the six pairwise comparisons here, all involve a dominance relationship. In the actual choice sets offered to the respondents five of the nine had no dominance relationships, two had three such relationships, one had two relationships and one had a single dominance relationship.

where V_j is the systematic component and ϵ_j is the unobserved, or random, component of total utility. The individual chooses alternative *j* if its total utility exceeds that of every other alternative in the choice set. Then the probability of choosing alternative j is

$$P(j)=P(U_j>U_s)\ \forall s\in C,\ j\neq s \tag{2}$$

where C is the specific choice set facing the individual. If the error terms are independently and identically distributed with a Weibull distribution, it can be shown that the probability of choosing alternative j may be written as the multinomial logit (McFadden, 1974)

$$P(j)=\frac{e^{X_j^{\prime}\beta/\mu}}{\sum_{s=1}^{J}e^{X_s^{\prime}\beta/\mu}} \tag{3}$$

where μ is a scale parameter. Note that as μ becomes large $P(j)\rightarrow 1/J$ and all alternatives are chosen with equal probability. As μ becomes very small the probability mass becomes concentrated in the choice with the highest systematic utility. Consequently the scale parameter may be interpreted as a measure of the error or lack of precision in the subject's choices. As noted earlier the scale parameter cannot be identified in the standard multinomial logit, so it is generally normalized to unity. Given this normalization, the vector of parameters β can be estimated by the method of maximum likelihood. Marginal willingness-to-pay for the k-th attribute can be estimated as $\mathrm{wtp}=\left.\frac{dx_t}{dx_k}\right|_{V=\bar{V}}=-\frac{\beta_k}{\beta_t}$, where *t* is the subscript of the payment attribute.

In this paper we allow the scale parameter to vary in two distinct ways: with cognitive ability and with the rank (first, second, or third) being analysed. The first way relates to the degree to which the respondents appear error prone. Recall that there were some nine pairwise comparisons where one of the pair dominated the other choice. Some individuals ranked some of these pairs in the ‘wrong’ order in the sense that the dominated choice was ranked as more preferred. It is worth recalling, however, that since the random utility model framework involves an error term, this ‘wrong’ ranking might simply be the result of an unusually large error term. However, it might alternatively indicate an inability to process the information, an attempt to fill in the questionnaire rankings without really evaluating the alternatives, or some other kind of behaviour not anticipated by the model. We attempt to determine which of these alternatives seems evident in our data. We do this by allowing the scale parameter to vary according to how many dominance violations each subject exhibits. We define up to 5 categories of cognitive difficulty (none, one, two, three and more than three dominance violation). We allow the scale parameter to vary across categories and normalize the error parameter for the first group to unity. The second way in which we allow the scale parameter to vary relates to data on second or third ranking. We allow these rankings to have different (and, we anticipate, larger) scale parameters because, in the ranking exercise, respondents were asked to turn their attention first to the most desirable alternative and only later to other alternatives. We anticipate that less care may be taken with lower choices and respondents may not so easily distinguish between them. For both these reasons, we think it appropriate to allow for the scale parameter to vary according to whether we are considering first, second, or third rankings.

Whether to allow the scale parameter to vary with ranking we view as being related to the efficiency issue of whether to use the information about second and third choices. Frequently in work in this area, the focus is on studying first choices versus the other possible choices. However, with ranked data, the fact that a respondent ranks a particular choice second rather than third or fourth contains all of the information necessary to estimate the conditional logit model without using first choices. However, using second and/or third choices may well introduce additional noise into the model -- especially since respondents were first asked to focus on the best alternative of the four and only subsequently to run to second and third choices. Using a variable scale parameter will be useful if it allows using this additional information and leads to reduced confidence intervals for our estimates.

Whether to allow the scale parameter to vary with apparent cognitive ability is a more complicated issue in our view. There are two possibilities that could give rise to these 'errors' in choosing dominated alternatives. On the one hand, the random utility model is predicated on there being a random component associated with each option. The option is chosen that has the highest random utility. Thus a large positive drawing could lead theoretically to a 'dominated' choice being chosen (or a large negative error component for a dominating choice being ranked below a dominated one). If one takes this view of the world, then the observed inconsistencies may just be the result of unusually large (positive or negative) error draws. An alternative view of the world is that errors made in choosing (or ranking higher) dominated choices are an indication that the respondent has difficulty processing the information presented to him/her. That is, the respondent is different from the normal respondents who did not make such errors and there is an underlying heterogeneity in respondents. Some are simply less able to do the task than others.

How can we distinguish between these alternative views of the world? Return to the first hypothesis for the moment. If there is an underlying homogeneity across individuals and the apparent errors observed are simply the result of some large random draws, then the other rankings (for other choice sets) that the individual makes should not be affected by such a large draw. It is this idea we will use to try to rule out the homogeneity hypothesis. In particular, if we look at the other choices of individuals who made 'cognitive errors', we should not notice any difference from the 'norm' in those choices. They should be no more likely to make errors in other choices than a randomly selected individual. To address this issue we will allow for the individuals who make some 'errors' to have a larger variance term (in all their choice sets) than other individuals in the sample. A further problem arises, here, however. If we simply identify individuals who have made some apparent errors and then allow for a larger variance term in those cases, surely we must find such a scale term significant because we are using the very observations with errors we are assuming to be unusually large. To address this concern, we will estimate the model allowing for larger variance terms with and without the observations that are assumed to be in 'error'. If we find a large estimated variance (or scale effect) terms when we have eliminated all the choice sets where there were errors (but keeping the other choices of individuals who have made errors) this will confirm an underlying heterogeneity in these respondents.

To show formally how we allow variable scale parameters, we restore the subscripts for individual and choice set and write the likelihood of any alternative j being assigned rank r in choice set n by individual i as

$$P_{in}(j,r)=\frac{e^{X_{inj}^{\prime}\beta/\mu_{inr}}}{\sum_{s=r}^{J}e^{X_{ins}^{\prime}\beta/\mu_{inr}}} \qquad \textbf{(4)}$$

We want to allow the scale parameter to vary with the rank (1^{st}, 2^{nd}, or 3^{rd}) and the cognitive category (the base category or no errors, plus 4 others). We accomplish this by setting $\mu_{inr}=\sigma^{C}{}_{c_i}\sigma^{R}{}_{r}$, where $\sigma_c^C,\ c=1,..,5$ .is the cognitive scale parameter (where c_i is the cognitive category for the *i*-th individual) and $\sigma_r^R,\ r=1,2,3$ is a ranking scale parameter and we adopt the normalization $\sigma_1^R=\sigma_1^C=1$. This approach economizes on the number of parameters by requiring that the error parameter for any observation be the product of the ranking scale parameter and the cognitive scale parameter.

The likelihood function to be maximized then becomes

$$L=\prod_{i\in I}\prod_{n\in N}\prod_{r\in R}P_{in}(j,r) \qquad (5)$$

Note that a subject's cognitive classification, σ_{c_i}, is endogenously determined. This can be seen by writing $\sigma_{c_i}=\sum_{c=1}^{5}\sigma_c D_i^c$ where D_i^c is a dummy variable indicating that the respondent who generated the observation has been placed in the c-th cognitive group. Suppose that choice j in equations (1) and (2) refers to a choice which is dominated by some other choice in the choice set. Even if there is no underlying heterogeneity among respondents, it is clear from equations (1) and

(2) that a high realization of the random variable e_j will lead to the choice set being flagged as containing a dominance violation.

As explained above, the estimates of willingness to pay are found by looking at the ratio of coefficients estimated in the conditional logits. These estimates are non-linear functions of the coefficients. We test for the significance of these estimates of willingness to pay by using a Wald test. However, we are also interested in comparing the willingness to pay estimates across various models and samples. For this we find it useful to think of the confidence limits surrounding the point estimates of willingness to pay. We calculate these by a simulation method that takes drawings from a multivariate normal distribution with mean and variance as given in the conditional logit estimates and calculates the willingness to pay measures 10,000 times. We then find the willingness to pay values of the bottom and top 2.5% and think of these as the 95% confidence limit

Results

First choices

We begin by analysing the information contained in first choices, focussing on the issue of cognitive ability. The results from the conditional logit estimates of first choices are presented in Table 1. The first column reports a baseline model (Model 1) in which all observations are used and the single scale parameter is normalized to unity. The coefficients on health (H), visibility (V), odour (O), fallout (F) and taxes (T) appear near the bottom of the Table. Note that in this and all the remaining columns of Table 1 (and subsequently in Table 2), all coefficients are significant at the 1% level when tested against a null of zero. The interpretation of these coefficients depends on the coding used in the study. Each of the attributes is coded here as [-1, 0, +1] with 0 being the status quo, -1 being a situation one-third worse than the one described as the status quo, and +1 being the situation one-third better than the status quo. Tax increments or decrements, on the other hand, are coded in the actual dollar changes. The coefficients on the attributes should thus be positive and the coefficient on the tax attribute negative and indeed they are. The magnitude of the coefficients indicates that the health effects (hospital admissions and deaths per month) have a much bigger impact on perceived utility than the other air quality attributes. Moreover, we can see from the other coefficients that individuals regard the one-third worsening of black fallout as being more of a concern than odour or visibility (in that order).

At the top of the column we report the willlingness-to-pay (WTP) estimates computed from the formula $\mathbf{wtp} = \left.\frac{dx_t}{dx_k}\right|_{V=\bar{V}} = -\frac{\beta_k}{\beta_t}$. Thus the WTP for a one-third health improvement (as defined in the survey) is calculated as 0.62/0.019 = 84.20. Individuals appear to be willing to pay just over $84 extra per household in taxes per month in order to reduce the hospital admissions and deaths by one-third. Similarly, they would be willing to pay about $20 per month for a reduction in the frequency of poor visibility (again by one-third), and so on.

The WTP estimates are highly statistically significant. Since WTP is a non-linear function of the coefficients, we adopt a Wald test which we conducted for the WTP estimates in column one (and all other columns) in Table 1. The null hypothesis that the WTP estimates are zero is soundly rejected in all cases (at considerably better than the 1% level). Again, because we so systematically reject the hypothesis we do not bother to flag the estimates to indicate significance. Because it will be useful in later discussions, however, we do report beneath each of these WTP estimates an empirical estimate of the 95% confidence range, that is the range which leaves 2½ % of the simulated values in each of the tails of the distribution. If these WTP estimates were normally distributed (or otherwise symmetrical) , which of course they are not, half the range would lie to either side of the point estimate[9]. Even though the distributions are not symmetric, we find the range a useful statistic for comparing alternative models.

Before leaving this first column it is worth mentioning the number of observations reported at the bottom of the column. Recall from the earlier discussion that 515 surveys were returned. How then did the number of observations get to be 15,516? First recall that each respondent was asked to rank 9 choice sets which, had all respondents completed all the rankings, would have given us 4635 rankings. The conditional logit treats each of the choices (4 of them) within a choice set as an observation, so if all surveys were complete, we should have had 18,540 observations. The 15516 reported is 84% of this number which means that 16% of the choice sets were left blank.

We turn now to the issue of cognitive ability and do so by focussing on the WTP estimates. To do so we compare Model 1 to Models 1A and 1B, in which we drop observations that appear to be associated with respondents' cognitive difficulty. These models are reported in columns two and three of Table 1. Model 1A drops only the **choice sets** in which a dominated choice was ranked as being preferred to a dominating choice (about 12% of the choice sets).[10] The willingnesses-to-pay fall by from 7% to 20% . More importantly, the estimated confidence

[9] In fact, they tend to be positively skewed with the upper bound further from the point estimate than the lower bound. For example, the range for WTP H is 71.4 to 102.7.

[10] We flag such observations with a binary variable, *derror,* coded 1 if a choice set contains a dominance violation. Retaining only cases where derror = 0 means dropping the choice sets in which an error was made. It is also possible to drop simply the cases where the dominance violation involved the first place ranking in the choice set (rather than those with an error in any of the rankings). Doing this leads to much the same conclusions as in Model 1A.

intervals are substantially reduced (by from 28% to 40%). Model 1B drops **all** the choice sets of **respondents** who ever made an 'error'[11] Slightly over 50% of the choice sets are dropped in this case. Here the willingness-to-pay estimates are slightly lower and the confidence intervals slightly wider than in Model 1A. These results suggest that the preferences of subjects exhibiting cognitive difficulties may vary systematically compared to those with no dominance errors. Although omitting observations from these subjects may appear to improve the efficiency of the estimates, it may also introduce a bias into the average willingness-to-pay estimates.

We next investigate whether introducing variable scale parameters into the model allows us to improve efficiency without introducing a potential bias. Consider Model 2, which is estimated on all the data but which introduces additional scale parameters reflecting the number of dominance violations exhibited by a subject. Moving down the Model 2 column (the fourth column of Table 1), we discover that the scale parameter for individuals making exactly one error (i.e. for whom domtot=1) is 1.20 times the scale parameter of the reference group, who exhibited no cognitive errors. The scale parameter for individuals making exactly 2 errors is 1.83 times the reference scale parameter and the scale parameters for subjects making 3 and more than 3 errors are 2.08 and 7.42 times the reference levels, respectively[12]. This steady increase in the scale parameter as apparent cognitive difficulty increases is entirely consistent with the reasoning

---

[11] *domtot* is a variable which records for each individual how many dominance 'errors' he or she made. Retaining only those cases with *domtot* = 0 drops all cases where the respondent ever made an 'error'.

[12] We experimented with various ways of categorizing the cognitive difficulties in our estimation. We found that the four scale parameters were the maximum we could estimate without running into difficulties in convergence in some cases. As well, likelihood tests indicated that two scale parameters (1 error and more than one error) was a preferred specification relative to one scale parameter, three was preferred to two, and four was preferred to three.

sketched above. All the scale parameters are significantly different from unity at conventional levels and a log likelihood test comparing Model 1 to Model 2 shows that the null of all the terms being equal to unity can be easily rejected. Beneath each estimated scale parameter is recorded the 95% asymptotic confidence interval. Note that for the most part the scale terms are increasing and the confidence intervals are non-overlapping.

Note further that the WTP estimates in Model 2 generally lie between those of Model 1 and Model 1A but much closer to Model 1A. The confidence intervals in Model 2 are slightly larger than those in Model 1A but generally smaller than those in Model 1B. Thus, correcting for cognitive errors by allowing scale parameters to vary provides similar information to dropping the offending 'error' observations. Model 2 has the advantage, however, of not arbitrarily discarding preference information embodied in these observations and possibly introducing bias into the estimates.

The significance of the scale parameters in Model 2 appears to confirm that respondents are heterogeneous in respect of their cognitive ability. As noted earlier, however, the classification of observations into cognitive difficulty groups is not really exogenous. We test for such spurious correlation by re-estimating Model 2 while dropping the observations where the actual 'errors' occurred. These results are reported as Model 2A. The estimated willingnesses-to-pay and confidence intervals are generally slightly smaller than in Model 2. On the null hypothesis of no heterogeneity of the subject pool, the scale parameters would be indistinguishable from unity. It is most reassuring to find that the scale parameters rise as in Model 2 and for the most part, the ranges are non-overlapping. Again, a log likelihood test comparing Model 1A to Model 2A shows that the null hypothesis that the scale parameters are all unity is easily rejected.[13] Moreover, the 95% confidence intervals all exclude unity (although in the case of a single error, not by much) and with one exception do not overlap. This is a strong confirmation that the cognitive classification has captured a significant aspect of individual's decision-making.

First, Second and Third Choices

We now investigate the role that variable scale parameters may play in combining information from different rankings. Table 2 presents the results of estimating variants of equation (5), using data from first, second and third choices. Models 3, 3A and 3B are parallel to Models 1, 1A and 1B in Table 1. The basic model (Model 3) is estimated on all observations. Model 3A is estimated after dropping all choice sets exhibiting a dominance error and Model 3B is estimated after dropping all observations (choice sets) for error prone respondents. No scale parameters are included in this model. The results are quite similar to those in Table 1. The coefficients rise as more (doubtfully valid) observations are eliminated, while the willingness to pay values fall and the 95% ranges for the WTP measures fall as well.[14] The main difference is that in moving from Model 1A to 1B the confidence intervals rose while here, moving from 3A to 3B , they fall. Relative to Model 1, Model 3 yields higher willingness-to-pay estimates and narrower confidence intervals. The reduced confidence intervals suggest that including the extra ranking information has improved the efficiency of the estimates. It is clear, however, that heterogeneity among respondents continues to affect the results. As in the analysis of first choices, excluding choice

[13] Twice the difference in log likelihoods is of the order of 226, while the chi-squared critical value (at 5%, 4 restrictions) is 9.50.

[14] As in the previous table, all the coefficients at the bottom of the table are highly significant with P values of zero to 3 significant digits. We avoid cluttering the tables with this information.

sets (Model 3A) or respondents (Model 3B) for which or for whom a dominance violation has been identified leads to reductions in the estimates of WTP.

Models 4, 5 and 6 introduce the variable scale parameters. Model 4 introduces scale parameters for second and third choices while Model 5 introduces them for cognitive difficulty (or error-proneness). Model 6 combines the scale parameters for second and third choices and for cognitive difficulty. Model 6A estimates Model 6 on the reduced data set which excludes all choice sets in which a dominance violation was made.

Turning first to Model 5, we discover that estimating the model with variable scale parameters for cognitive difficulty on all the data yields results quite similar to those obtained by estimating the model without scale parameters while dropping choice sets with dominance violations (much closer to Models 3A or 3B than to Model 3). Both the estimated willingnesses-to-pay and the range of the confidence intervals fall to approximately the same level as estimated in Model 3A, however the WTPs are slightly higher and the confidence intervals slightly narrower in Model 4. The scale parameters are highly significant on a log-likelihood test. Moreover the point estimates of the cognitive scale parameters themselves are all greater than unity and increasing in value as the number of cognitive errors increases. The first confidence interval does not incorporate unity and the remaining intervals are all non-overlapping. Thus the scale parameters behave in exactly the manner expected. Overall, Model 5 strongly suggests that efficiency is improved by incorporating both ranking information and variable cognitive scale parameters.

The story is somewhat different if variable scale parameters are used only to reflect heterogeneity across ranking of first, second or third choices. Model 4 is comparable to Model 3 but includes variable scale parameters for the ranking being considered (second choices or third choices). The scale parameters are jointly significant[15] and they do increase with the choice rank being analysed, yet they do not greatly influence the WTP estimates nor the confidence intervals. Indeed both are somewhat greater in Model 4 than in Model 3. Moreover, since the σ_r^R do not suffer from the possibility of endogenous in the way the cognitive scale parameters do, there is no problem of endogeneity bias in this case.

Models 6 and 6A incorporate both types of scale parameter – the σ_r^R for second and third choices and the $\sigma_{c_i}^C$ for cognitive difficulty. Comparing the log likelihood from Model 6 with those in Models 3, 4 or 5, we see a substantial reduction in the likelihood value and a clear rejection of the null that the cognitive and ranking scale parameters are all unity, either collectively or separately. This test is not completely valid, however, because of the endogeneity of the cognitive scale parameters. To test whether the significance of the scale parameters is spurious we estimate Model 6A, which differs from Model 6 only in dropping the observations (choice sets) where ranking violations occurred. Comparison of the log-likelihoods of Models 3A and 6A shows that the scale parameters are strongly and significantly different from unity.[16] Models 6 and 6A exhibit patterns we have already seen: third choices have a bigger scale term than second

[15] To see that the two Scale Terms are jointly significant, notice that twice the difference in log likelihood values between Model 3 and Model 4 is well over 300. This statistic is distributed Chi Squared with 2 degrees of freedom and has a critical value of about 6.0.

[16] We have not been able to discover why STATA's maximum likelihood routine drops one extra observation in Model 3A as compared to Model 6A. STATA drops observations if no choice is uniquely ranked best but normally it would drop all the observations involved in that ranking. If the problem were in a third choice situation, then there should still be two observations involved and dropped.

choices and the ranges are not overlapping, the cognitive difficulty terms rise in value as you go down the columns and the 95% confidence limits are not overlapping. Finally the WTP estimates are smaller than in Model 3 or 4 and smallest in Model 6A.

The value of analysing second and third rankings is seen by comparing Model 2A (from Table 1) with Model 6A from Table 2. Including second and third choices with appropriate scale parameters more than doubles the effective number of observations and yields **higher** willingness-to-pay estimates and **narrower** confidence intervals than those obtained by analysing first choices alone, even when action is taken to correct for the endogeneity bias introduced by assigning cognitive categories on the basis of internal evidence.

Conclusions

This paper has investigated the value of using variable scale parameters to account for heterogeneity in responses to willingness-to-pay surveys. Specifically, we considered heterogeneity in cognitive ability across respondents and the potential deterioration of precision in ranking second and third choices. Heterogeneity in cognitive ability was explored through internal consistency checks in the survey data. Cognitive scale parameters were clearly, significantly and positively related to the frequency with which respondents ranked dominated choices ahead of undominated ones. This significance was retained even after corrections designed to eliminate endogeneity bias. Ignoring heterogeneity in the cognitive ability of respondents biases willingness-to-pay estimates upwards. Accounting for it by introducing variable scale parameters and by discarding offending observations leads to comparable results, but the variable scale parameter method has the advantage of maintaining the entire dataset.

We have shown that consideration of second and third choices can increase the efficiency of willingness-to-pay estimates, as shown by the reduced range of the confidence intervals in models that incorporate all rankings (first, second, and third). Variable ranking scale parameters introduced to account for heterogeneity in responses to different rankings are statistically significant, but they do not improve the efficiency of the estimates either by themselves (compare Model 4 to Model 3) or in combination with cognitive scale parameters (compare Model 6 to Model 5). With this data set, the WTP with generally smallest confidence intervals are achieved by exploiting multiple ranking information while introducing variable scale parameters only for identified heterogeneity in cognitive ability (Models 5). To ensure a conservative estimate, it appears best to drop the choice sets used to identify dominance violations.

TABLE 1: FIRST CHOICES Estimates of the Model with and without allowance for error terms

	Model 1	**Model 1A**	**Model 1B**	**Model 2**	**Model 2A**
	all obs	derror=0	domtot=0	all obs	derror=0
WPT H	84.2	67	69.6	73.9	68.6
95% range	31.3	18.8	26.8	23.5	19.8
WPT V	19.9	18.6	14.8	18.2	17.9
95% range	8.8	6.3	7.7	7.1	6.4
WPT O	28.4	22.4	20.5	23.5	21.3
95% range	12.5	8.0	10.4	9.3	8.0
WPT F	31.1	27.2	26.3	28.5	27.7
95% range	10.6	7.0	8.7	8.1	7.2
SCALE TERMS					
one error				1.20	1.12
				(1.08-1.32)	(1.01-1.24)
two errors				1.83	1.61
				(1.60-2.06)	(1.40-1.84)
three errors				2.08	1.75
				(1.80-2.34)	(1.49-2.01)
more than 3 errors				7.42	3.86
				(3.97-10.88)	(2.44-5.29)
COEFFICIENTS					
Coefficient on H	1.62	1.86	2.38	2.34	2.34
Coefficient on V	0.38	0.52	0.51	0.57	0.61
Coefficient on O	0.55	0.62	0.70	0.74	0.73
Coefficient on F	0.60	0.76	0.90	0.90	0.95
Coefficient on T	-0.019	-0.028	-0.034	-0.032	-0.034
Number of Obs	15516	13648	7480	15516	13648
Log Likelihood	-3209.0	-2533.0	-1032.5	-2953.2	-2420.4

Model 1: Baseline model: first choices only

An A indicates choice sets are dropped where 'errors' are made
A B indicates choice sets are dropped for individual who make any errors

Model 2: This is the Baseline model with variance terms "error-proneness"
An A indicates choice sets are dropped where 'errors' are made

TABLE 2: Estimates of the Model with and without allowance for error terms: All choices

	Model 3	Model 3A	Model 3B	Model 4	Model 5	Model 6	Model 6A
	all obs	derror=0	domtot=0	all obs	all obs	all obs	derror=0
WPT H	99.4	84.2	80.1	101.1	86.7	86.5	80.4
95% range	19.9	13.2	12.5	25.2	12.8	15.4	13.1
WPT V	23.8	22.6	17.5	27.4	20.1	22.6	22.4
95% range	6.9	5.2	4.8	8.4	4.8	5.6	5.1
WPT O	30.2	27.4	23.0	31.3	25.8	26.0	24.2
95% range	8.2	6.1	5.6	9.9	5.6	6.5	5.7
WPT F	33.6	31.7	28.2	33.2	30.8	30.5	29.9
95% range	7.8	5.9	5.7	8.8	5.6	6.0	5.4
SCALE TERMS							
Second choice				1.38		1.24	1.17
				1.28-1.48		1.15-1.34	1.09-1.26
Third Choice				2.29		2.06	2.04
				2.03-2.55		1.85-2.27	1.83-2.25
one error					1.28	1.29	1.19
					1.18-1.38	1.19-1.39	1.09-1.29
two errors					2.07	2.09	1.74
					1.85-2.29	1.87-2.31	1.55-1.93
three errors					3.19	2.91	2.43
					2.75-3.63	2.54-3.27	2.09-2.77
more than 3 errors					11.66	10.24	4.86
					5.12-18.21	5.35-15.13	3.19-6.53
COEFFICIENTS							
Coefficient on H	1.31	1.54	2.08	1.63	2.06	2.50	2.44
Coefficient on V	0.32	0.41	0.45	0.44	0.48	0.65	0.68
Coefficient on O	0.40	0.50	0.60	0.50	0.61	0.75	0.74
Coefficient on F	0.44	0.58	0.73	0.53	0.73	0.88	0.91
Coefficient on T	-0.013	-0.018	-0.026	-0.016	-0.024	-0.029	-0.030
Number of Obs	34583	30416	16673	34583	34583	34583	30417
Log Likelihood	-9397.7	-7707.5	-3523.3	-9230.2	-8822.0	-8688.7	-7297.8

References

Adamowicz W., J.J. Louviere , and M. Williams (1994) “Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities.” *Journal of Environmental Economics and Management*,26:271-92

Chapman, R.G., and R. Staelin (1982) “Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model.” *Journal of Marketing Research* 19, 282-301

Diener, Alan. (1999) *Valuing Health and Air Quality Using Stated Preference Methods.* Ph.D. thesis, McMaster University.

Henscher, David; Jordan Louvière, and Joffre Swait (1999) “Combining Sources of Preference Data.” *Journal of Econometrics* 89, 197-221.

Johnson, F. Reed, Melissa Ruby Banzhaf, and William H. Desvousges (2000). Willingness to Pay for Improved Resiratory and Cariovasular Health: A Multiple-Format, Stated-Preference Approach. *Health Economics* 9: 295-317.

Krinsky, I. and A.L. Robb (1991). “Three Methods for Calculating the Statistical Properties of Elasticities: A Comparison.” *Empirical Economics,* 16; 1-11

Layton, David F. (2000) “Random Coefficient Models for Stated Preference Surveys”. *Journal of Environmental Economics and Management.* 40(1), 21-36.

Louviere J.J. (1987) “Conjoint Analysis Modelling of Stated Preferences.” *Journal of Transport Economics and Policy*, 22:93-118

Louviere, J.J. (1988) *Analyzing Decision Making: Metric Conjoint Analysis*. California, USA:Sage Publications

McFadden D. (1974) “Conditional Logit Analysis of Qualitative Choice Behavior”. In Zarembka P, editor. *Frontiers in Econometrics*. New York: Academic Press. 1974; p.105-42

Petersen, R.G. (1985) *Design and Analysis of Experiments*. New York: Dekker

Ryan, Mandy, Emma McIntosh, and Phil Shackley (1998) Methodological Issues in the Application of Conjoint Analysis in Health Care. *Health Economics* 7(4), 373-78..

Schulze W., G. McClelland, M. Doane, E. Balistreri, R. Boyce, B. Hurd, and R. Simenauer (1995) *An Analysis of Stated Preferences for Superfund Site Cleanup*. Department of Agricultural, Resource, and Managerial Economics, Cornell University, unpublished paper.

Figure 1: Sample Question Taken from the Questionnaire

	State of the Environment			
Attribute	Choice A (Current situation)	Choice B	Choice C	Choice D
Bad Odour	4 days of bad odour per month	One-third Worse	One-third Better	Same
Black Fallout (BFO)	3 days of BFO per month	One-third Worse	One-third Better	Same
Poor Visibility	3 days of poor visibility per month	One-third Worse	One-third Better	Same
Health Effects	18 extra hospital admissions per month and 2 extra deaths per month, compared to perfectly clean air	One-third Worse	One-third Better	Same
Monthly Property Taxes or Rent	$200 per month for typical household	$5 more taxes/rent	$5 less taxes/rent	$5 more taxes/rent
Rank	*2*	*4*	*1*	*3*

Please put a 1 under the choice you like BEST.
Please put a 4 under the choice you like LEAST.
Consider the remaining choices. Put a 2 under the one you like best. Put a 3 under the other one.

APPENDIX

** STATA CODE (do file) to estimate Conditional Logit Model with scale parameters by ML

```
set mem 16m
 capture program drop mlclgt6
 program define mlclgt6
    args todo b lnf
    tempvar theta1     /*  X_j beta  - inner product of parameters and j-th observation */
    tempname sig1 sig2 sig3 sig4 sig5 sig6 /* error parameters*/
    tempvar rr         /*  log of ratio of numer/denom */
    tempvar denom      /* denominator of the density */
    tempvar numer      /* numerator of the density f(x_i) */

    mleval `theta1' = `b' , eq(1)  /* creates X'Beta*/
    mleval `sig1' = `b' , eq(2) scalar /* sets up variable for scale parametrs*/
    mleval `sig2' = `b' , eq(3) scalar /* sets up variable for scale parametrs*/
    mleval `sig3' = `b' , eq(4) scalar /* sets up variable for scale parametrs*/
    mleval `sig4' = `b' , eq(5) scalar /* sets up variable for scale parametrs*/
    mleval `sig5' = `b' , eq(6) scalar /* sets up variable for scale parametrs*/
    mleval `sig6' = `b' , eq(7) scalar /* sets up variable for scale parametrs*/

** Data is arranged in groups with identifier given by 'grp'
** 'grp' is unique for each group of observations
** a single choice set has 4 lines of data for first choice, 3 for second and 2 for third
** 'grp' has a different value for second choices than for first choices and thirds

**begin by treating every row of data as potentially being in the numerator

   gen double `numer' = `theta1'

**add scales terms for the various groups and 'errors'

    quietly replace `numer' = (`numer'/`sig1') if second == 1
    quietly replace `numer' = (`numer'/`sig2') if third  == 1
    quietly replace `numer' = (`numer'/`sig3') if domtot == 1
    quietly replace `numer' = (`numer'/`sig4') if domtot == 2
    quietly replace `numer' = (`numer'/`sig5') if domtot == 3
    quietly replace `numer' = (`numer'/`sig6') if domtot  > 3

    quietly replace `numer' = exp(`numer')
**now create a denominator for each line of data
    egen double `denom' = sum(`numer'), by(grp)
**now create the ratio
    g double `rr' = ln(`numer') - ln(`denom')
**use the stat code to form the likelihood by picking off the rows in which chosen = 1
    mlsum `lnf' = `rr'  if $ML_y1==1
    if `lnf' == . {exit}
  end
```

```
clear
use alldat

ml model d0 mlclgt6 (chosen = h v o f t, nocons)  /sigma1 /sigma2 /sigma3 /sigma4 /sigma5 /sigma6
ml init h = 2.5 hpos=2.5 hneg=2.5 v=.5 vpos=.5 vneg=.5 o=.5 opos=.5 oneg =.5 f=1 fneg = 1 fpos=1 t=-.04 tneg=
-.04 tpos=-.04 /sigma1 =1.5  /sigma2 =1.5 /sigma3 =1.5 /sigma4 = 1.5 /sigma5 = 1.5 /sigma6 = 1.5, skip
ml search
ml maximize
matrix temp = e(b)
matrix V = e(V)
matrix wtp = -temp/temp[1,5]
matrix list wtp
clear
```

Recent working papers

2001-01	John B Burbidge	'Awkward Moments in Teaching Public Finance' John F Graham Lecture at the Atlantic Canada Economics Association Annual Meetings
2000-06	Alok Johri Lohn Leach	Middlemen and the Allocation of Heterogeneous Goods
2000-05	John B Burbidge Gordon M Myers	Tariff Wars and Trade Deals with Costly Government
2000-04	Donald Dawson Lonnie Magee	The National Hockey League Entry Draft, 1969-1995: An Application of a Weighted Pool-Adjacent-Violators Algorithm
2000-03	Neil Buckley Kenneth S. Chan James Chowhan Stuart Mestelman Mohamed Shehata	Value Orientations, Income and Displacement Effects, and Voluntary Contributions
2000-02	R Cooper Alok Johri	Learning by Doing and Aggregate Fluctuations
2000-01	Yorgos Y. Papageorgiou David Pines	Externalities, Indivisibility, Nonreplicability and Agglomeration